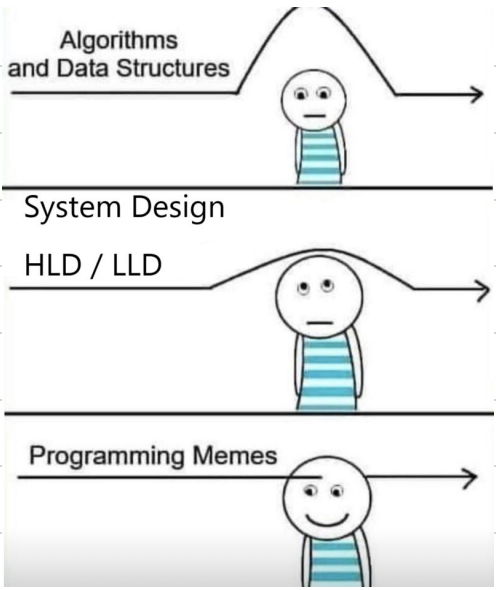
Session-10

ML-System Design-2

Netflix Recommendation Engine

Oct 29, 2025



Agenda

- Overview of Netflix Recommendation system-design

Difference between system design for machine learning engineers and software developers

Aspect	SDE	ML
Primary goal	Reliable feature delivery & Request/Response correctness	Accurate prediction + stable data/feature pipeline
Artifacts	API, Service, DB schema	pipeline, models, Feature, training code
Variability	API Req./Response, validate if of API ena	DRIFT
Metrics	Latency, throughput,	Offline/online metrics
Scale.	Stateless DBJ, horizontal (Ver	Raw/Curated data model
Failure	Service Crash, # of retries	Bad feature, drift

Where do we start ??

MLOps

1. Data Version
2. Model Structure
3. Training – Inference Skew
4. Drift

System Design

1. LLD -> Low Level Design
   - Implementation details
2. HLD -> High Level Design
   - architecture details

Design a Video Streaming Platform like Netflix

CLARIFYING QUESTIONS

1. 1 Billion Users — 200 million DAU
2. 1 million videos on platform
3. # of videos to be uploaded -> 1000
4. How many times user accesses netflix
   - 5

$$\frac{200 \times 10^6 \times 5}{24 \times 60 \times 60} = 12K \text{ req. / Second.}$$

2K / min

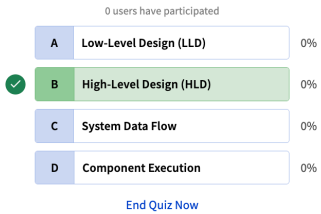
1 Functional Requirements

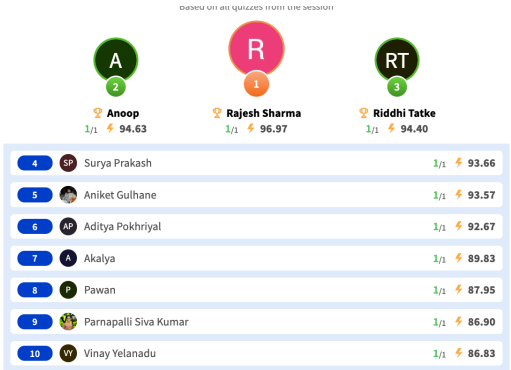
a. Users should be able to stream video on the platform (pause/resume video playback).
b. Personalized Recommendation, and non-personalized recommendation.
c. People should be able to search videos on the platform.
d. People should be able to rate/comment/like/dislike on the videos.
e. Diff video quality options.
f. Producers, directors, and teams should be able to upload video on the platform.
g. Should be available on platforms like iOS, Android, and Web.

2 Non-Functional Req.

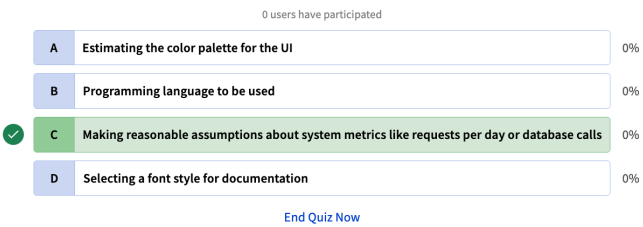
1. U. low latency & high availability.
2. Scalable & Efficient
3. Slick UI / smooth interact.

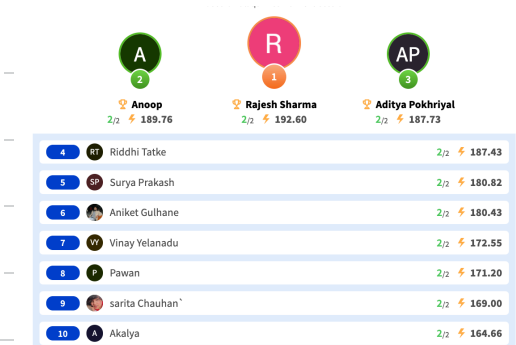
In the context of system design, what refers to the overall view of the system, including architecture, applications, and database management?





What is an essential aspect of system design in regards to understanding its scale?





Different Components in our MVP

1. Data Sources.
2. Data Ingestion Pipeline
3. Data Storage Systems.
4. Data Processing and Feature Engineering.
5. Recommendation Engine.
6. API Layer.
7. Caching
8. Monitoring and Logging.
9. Security and Compliance.

1 Data Sources

a Users / User-item Interaction

(i) Like / dislike / reviews / ratings / comments
(ii) Watch history
(iii) Age, gender, demographic info
(iv) User interaction with the platform
(v) Wish list / watch list
(vi) Daily time spent

b Content / Content metadata

1. Movies / Shows - (Genre, director, title, subtitle, Lang, cast, duration, descrip, Rating, KIDS - 18+)
2. Technical specification → 1080 / 4K / 480p / 360p
3. Release Year
4. Content popularity metric - IMDB - Rotten tomato

(5) External data — (1) Social media trend, sentiment

(2) Famous third party reviews.

Users:
- User ID
- User Name
- User Email
- Address (Can have Communications and Preferences Tables)
- Phone Number
- Password
- Joined Date
- Plan ID
- Is Active
- Payment ID

Sessions:
- ID
- User ID
- Show ID
- Device ID
- Playback Start Timestamp
- Playback End Timestamp
- Paused Time
- Session Time
- Pause Time

Devices:
- Device ID
- Device Type
- Is Active

Genre:
- Genre ID
- Genre Type
- Is Active

Plans:
- Plan ID
- Plan Name
- Streaming Limit
- Plan Base Rate
- Plan Start Date
- s Active

Payments:
- Payment ID
- User ID
- Payment Date
- Payment Method
- Total Amount
- Transaction ID
- Payment Status

Shows:
- Show ID
- Show Desc
- Show Type(Movie , TV Show , Documentary)
- Genre ID
- Length
- Release Date
- Is Downloadable
- If Subtitles Available (Y/N)
- Audio Desc Available (Y/N)

2 DATA INGESTION PIPELINE

CLIENT

1. Click stream data
2. Information regarding device usage.
3. Wakeup / shutdown ups

SERVER

1. Recommendation results
2. Which all hotels were clicked & review


```
"userId": "user123",
"sessionId": "session456",
"events": [
{
"eventType": "view",
"page": "/home",
"timeSpentSeconds": 10,
"timestamp": "2025-01-27T10:00:00Z"
},
{
"eventType": "scroll",
"scrollPosition": 300,
"timestamp": "2025-01-27T10:00:05Z",
"page": "/home"
},
{
"eventType": "click",
"elementId": "signup-button",
"timestamp": "2025-01-27T10:00:10Z",
"page": "/home"
},
{
"eventType": "view",
"page": "/product/123",
"timeSpentSeconds": 30,
"timestamp": "2025-01-27T10:01:00Z"
},
{
"eventType": "form_submission",
"formId": "contact-form",
"fields": {
"name": "John Doe",
"email": "john.doe@example.com"
},
"timestamp": "2025-01-27T10:02:00Z",
"page": "/contact"
}
```

3 Non- Structu / Structu - Data Storag

```
// Electronics Product
{
"product_id": 1,
"name": "Smartphone X",
"category": "Electronics",
"price": 999,
"attributes": {
"batteryLife": "24h",
"warrantyPeriod": "2 years",
"brand": "TechBrand"
},
"reviews": [
{
"user_id": 101,
"rating": 5,
"comment": "Excellent phone with great battery life!"
},
{
"user_id": 102,
"rating": 4,
"comment": "Good performance but a bit pricey."
}
]
}
// Clothing Product
{
"product_id": 2,
"name": "T-Shirt Y",
"category": "Clothing",
"price": 29,
"attributes": {
"size": "L",
"material": "Cotton",
"gender": "Unisex"
},
"reviews": [
{
"user_id": 103,
"rating": 4,
"comment": "Comfortable and fits well."
}
]
}
// Books Product
{
"product_id": 3,
"name": "Novel Z",
"category": "Books",
"price": 15,
"attributes": {
"author": "Author A",
"ISBN": "123-4567890",
"genre": "Fiction"
},
"reviews": []
}
```

99.90% !

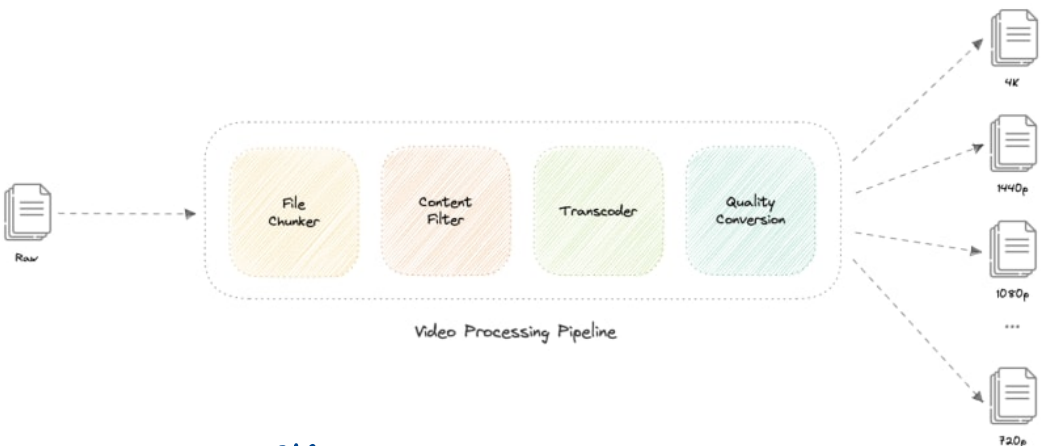
1. Streaming
2. Downloadin

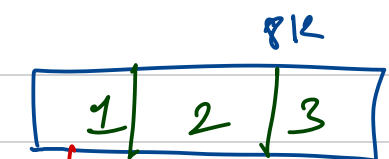
What is the main difference between horizontal scaling and vertical scaling?

0 users have participated

A	Horizontal scaling adds more machines, while vertical scaling adds new software.	0%
B	Horizontal scaling adds physical locations, while vertical scaling focuses on security enhancements.	0%
C	Horizontal scaling adds resources to a single machine, while vertical scaling adds more machines.	0%
✓ D	Horizontal scaling adds more machines, while vertical scaling adds resources to a single machine.	0%

End Quiz Now





8K / 4K / 2K / 1K (720P / 480P

480P

1080P

480P

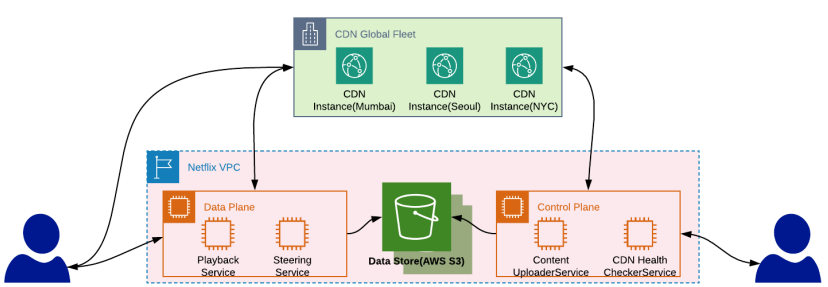
Video Resolution	Video Bitrate
240p	300 kbps
360p	500 kbps
HD 480p	1000 kbps
HD 720p	1500 kbps
HD 720p	2250 kbps
Full HD 1080p	3000 kbps
Full HD 1080p	4500 kbps
Quad HD 1440p	6000 kbps
Quad HD 1440p	9000 kbps
4K UHD 2160p	13,000 kbps
4K UHD 2160p	20,000 kbps

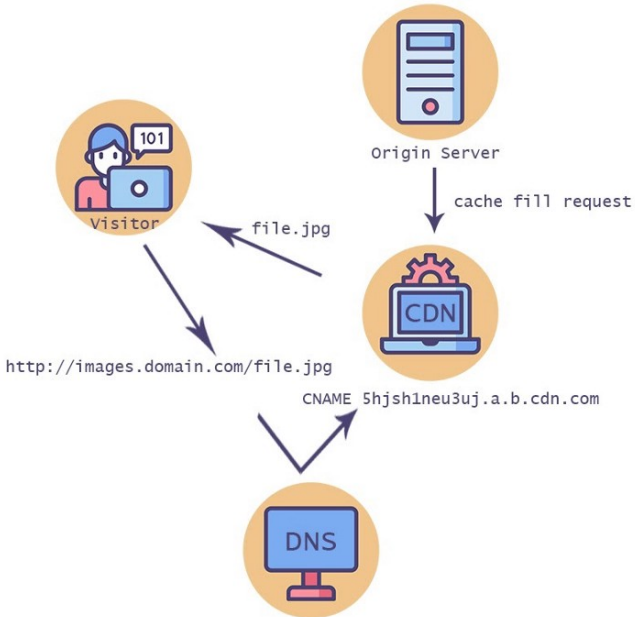
VIDEO STREAMING.

CDN

Server

Flowchart



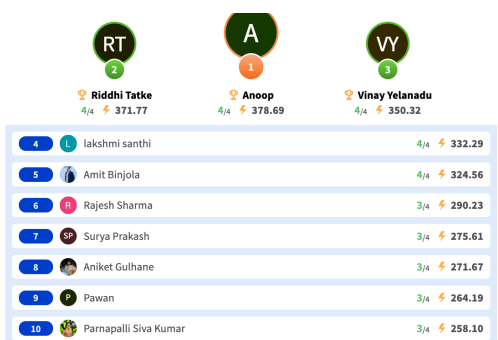


What is the role of a Content Delivery Network (CDN) in video streaming?

0 users have participated

A	To slow down the delivery of content to manage server load	0%
B	To provide a single centralized server for content storage	0%
✓ C	To offer geographically distributed servers for fast content delivery	0%
D	To increase internet connection speeds universally	0%

End Quiz Now



https://x.com/_trish_xD/status/1885662310091325737

Which component of Netflix's recommendation system uses an item-item similarity matrix to suggest content?

0 users have participated

A	Trending Now algorithm	0%
B	Award-Winning Comedies selector	0%
✓ C	Video-Video Similarity Ranker a.k.a Because You Watched (BYW)	0%
D	Continue Watching functionality	0%

End Quiz Now